Reading HDF5

Getting and Cleaning Data - Week 2

Krystella

2/11/2022

HDF5 - Hierarchical Data Format

- used for storing large data sets and structured data sets
- Stores a wide range of data types
- Data is stored in zero or more groups along with their metadata

R HDF5 Package

```r
knitr::opts_chunk$set(echo = TRUE, eval = FALSE, results = "hide")
```

```r
# Installing and loading

if (!require("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
BiocManager::install()

library(BiocManager)

BiocManager::install("rhdf5")
library(rhdf5)

created = h5createFile("example.h5")
created
```

- Reference: http://www.bioconductor.org/packages/release/bioc/vignettes/rhdf5/inst/doc/rhdf5.pdf - Link no longer available but search rhdf5 tutorial on site

Create Groups

- Once we create the rhdf5 file, we can create groups within the file

```r
created = h5createGroup("example.h5", "foo")
created = h5createGroup("example.h5", "baa")
created = h5createGroup("example.h5", "foo/foobaa")
```

```
  # creating a **subgroup** of the group "foo" called "foobaa"

h5ls("example.h5")
```

Writing to Specific Groups

```
# Create a matrix, A, and write that matrix to a particular group using the
# command **h5write**, where "example.h5" is the file and "foo/A" is the group
# within that file

A = matrix(1:10, nr=5, nc=2)
h5write(A,"example.h5","foo/A")

# Create a multidimensional array, B, and we can again use the command
# **h5write** to write the array to a particular group

B = array(seq(0.1,2.0,by=0.1), dim=c(5,2,2))
attr(B, "scale") <- "liter"
h5write(B, "example.h5", "foo/foobaa/B")
h5ls("example.h5")
```

Writing a Data Set Directly

```
df = data.frame(1L:5L,seq(0,1,length.out=5), c("ab", "cde", "fghi", "a", "s"),
                stringsAsFactors=FALSE)
h5write(df, "example.h5", "df")
h5ls("example.h5")
```

Reading Data

```
# Read data using the **h5read** command

readA = h5read("example.h5", "foo/A")
readB = h5read("example.h5", "foo/foobaa/B")
readdf = h5read("example.h5", "df")

readA
```

Reading and Writing Chunks

```
h5write(c(12,13,14), "example.h5", "foo/A", index=list(1:3,1))

  # Write the values (12,13,14) to the first three rows of the first column
  # (index=list(1:3,1)), of the dataset "foo/A", in the file "example.h5"

h5read("example.h5", "foo/A")
```